

AD-A046 502 CALIFORNIA UNIV DAVIS DEPT OF MATHEMATICS
ON THE POWER OF THE X SQUARED GOODNESS OF FIT TEST AT SIGNAL PL--ETC(U)
SEP 77 L D KAISER , F J SAMANIEGO AFOSR-77-3180

UNCLASSIFIED

TR-4

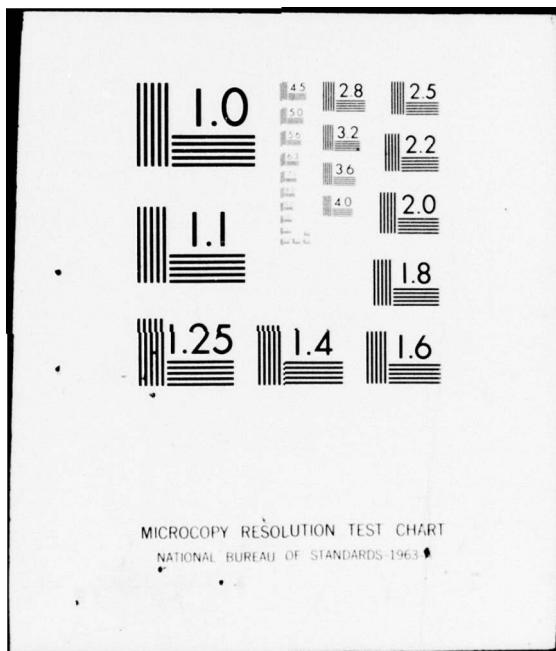
AFOSR-TR-77-1256

NL

F/G 12/1

| OF |
AD
A046502



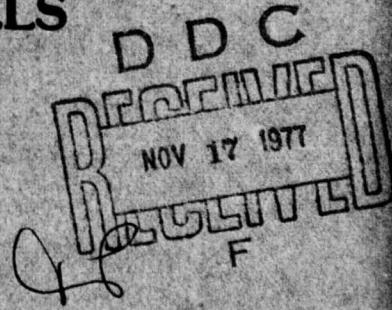


AFOSR-TR-77-1256

AD A046502

TECHNICAL REPORT SERIES
ON
SIGNAL PLUS NOISE MODELS

Approved for public release;
distribution unlimited.



Department of Mathematics
University of California
Davis, California



FILE COPY

**On the Power of the χ^2 Goodness of Fit
Test at Signal Plus Noise Alternatives**

Lee D. Kaiser and Francisco J. Samaniego

Technical Report No. 4

September, 1977

**Department of Mathematics
University of California
Davis, California**

DDC
REF FILE NOV 17 1977
RECEIVED F

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF TRANSMITTAL TO DDC
This technical report has been reviewed and is
approved for public release IAW AFR 190-12 (7b).
Distribution is unlimited.
A. D. ELOSE
Technical Information Officer

This research was supported in part by the Air Force Office of Scientific Research under Contract AFOSR-77-3180.

SUMMARY

ACCESS NUMBER	W-16 Section	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	B-17 Section	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
NTIS	UNARMED			
DDC	JUL 1 1971			
	BY	DISTRIBUTION AND USE NOTES		
		CLAL		
		<i>A</i>		

2

The χ^2 goodness of fit test is widely used as a means of validation of a probability model. Acceptance of a probability model on the basis of a χ^2 test is a dangerous procedure in the absence of a thorough knowledge of the power of the test for a broad class of reasonable alternatives. The power of the χ^2 test is examined via simulation at signal plus noise alternatives to the geometric and the Poisson distribution, and the power is seen to compare unfavorably with that of the likelihood ratio test. Likelihood ratio tests are thus advanced as goodness of fit criteria when signal plus noise alternatives are deemed relevant.

I. INTRODUCTION

The χ^2 goodness of fit test is by far the most widely used statistical technique for investigating the "adequacy" of a probability model in describing an observable random phenomenon. The procedure has been subjected to a considerable amount of quite justifiable criticism. Among the arguments against the procedure is the simple observation that no probability model constitutes an exact description of a stochastic process, and thus the null hypothesis that a probability model describes a process is known a priori to be false. Moreover, given a sufficiently large sample, any null hypothesis that a fixed probability model holds will be rejected with high probability. Thus, for large samples, departures from a probability model that are of no practical significance may result in a strong rejection of the hypothesis that this probability model describes the random process. Be that as it may, there remains an anomaly of even greater concern. The most common use of the χ^2 goodness of fit test is

the validation rather than the rejection of a given family of probability distributions. In other words, it is common practice to interpret the acceptance of a null hypothesis as a validation of a probability model, noting that such acceptance has shown that the data is not inconsistent with the model being entertained. It is a giant step indeed to the analysis that follows, in which the probability model is tacitly assumed to hold. The information necessary to justify this step is rarely available. The crucial information consists of a thorough knowledge of the power of the χ^2 test at a fairly broad set of reasonable alternatives to the null model. In the absence of such information, the experimenter remains unsure about the proper interpretation of acceptance, and may be led to accept a model only because the test procedure employed is unable to detect certain types of departures from this model.

Another peculiarity of the χ^2 goodness of fit test is the fact that a variety of substantially different probability models may be accepted on the basis of the test. Crow and Bardwell (1965) examined the fit of a collection of discrete probability models to several famous data sets. Models investigated include the Poisson, Neyman type A, Poisson-binomial, hyper-Poisson and Charlier type B series distributions. The data obtained by Rutherford and Geiger on the number of alpha particles emitted by a bar of polonium were fit by these distributions, and it was shown that the Poisson, type B and hyper-Poisson all fit well, that is, had χ^2 statistics with P-values around .20. We have fit a convoluted Poisson distribution to this data, the convolution of Poisson and Bernoulli distributions, and obtained a χ^2 statistic with a P-value in the same neighborhood. The utility of performing a χ^2 test might reasonably be questioned in the light of such ambiguous results.

We propose in this paper to investigate the power of the χ^2 goodness of fit test at particular signal plus noise alternatives. Since the χ^2 test is not directional, that is, is not designed with particular alternatives in mind, one would expect that there are more powerful tests available. For the two families of signal plus noise distributions investigated in this paper, we have estimated the power of likelihood ratio tests for comparison purposes. As predicted, the likelihood ratio procedure is superior, sometimes strikingly so.

Tests for detecting the presence of signal plus noise distributions have received attention in a recent paper by Sclove (1977). In that paper, tests are proposed for the hypothesis of a Poisson-Normal convolution against the broad alternative of an infinitely divisible distribution. A test is also proposed for the hypothesis of normality against the alternative of a Poisson-Normal convolution. These tests are based on moment estimators of cumulants of the underlying distribution, and the power of these tests remains to be investigated.

In the next section, we derive some results concerning the maximum likelihood estimates of parameters of the signal plus noise distributions studied here. These estimates are later used in computing the likelihood ratio statistic. In particular, we derive the maximum likelihood estimates of the parameters of the geometric-Bernoulli convolution and establish a relation between the MLE's of the parameters of the Poisson-Bernoulli convolution which facilitates the numerical search for the MLE's. In Section III, we describe the procedures used in the generation of random samples from these distributions, and make remarks on the numerical methods used. Tables summarizing our simulation study are given in Section IV,

followed by a discussion of the simulation and some concluding remarks.

II. MAXIMUM LIKELIHOOD ESTIMATION

Two signal plus noise distributions are studied in this paper:

(1) The distribution of the sum $X = Y + Z$ of independent variables Y and Z where Y has a geometric distribution with parameter π and Z is a Bernoulli variable with parameter p . The probability mass function of the variable X is given by

$$P(X=x|\pi, p) = \begin{cases} (1-\pi)(1-p) & \text{if } x = 0 \\ (1-\pi)((1-p)\pi+p)\pi^{x-1} & \text{if } x=1,2,\dots \\ 0 & \text{otherwise.} \end{cases}$$

(2) The distribution of the sum $X = Y + Z$ of independent variables where Y has a Poisson distribution with parameter θ and Z is Bernoulli with parameter p . The probability mass function of this variable is given by

$$P(X=x|\theta, p) = \frac{e^{-\theta}\theta^{x-1}}{x!} ((1-p)\theta+px), \quad x=0,1,2,\dots$$

Maximum likelihood estimation for samples from either of these distributions when the Bernoulli parameter p is known is a simple application of work in Samaniego (1976) and Samaniego (1977). We address below the problem of maximum likelihood estimation of the parameter pairs in these two families of distributions. Prior to our derivations in this regard, we examine the question of identifiability of these two-parameter families. Sclove and Van Ryzin (1969) obtained moment estimators for a variety of signal plus noise distributions, including the Poisson-binomial convolution, and established

the identifiability of this distribution in that paper. We therefore turn our attention to the geometric-Bernoulli convolution, which can be shown to be identifiable as follows. If X is distributed as the sum of independent geometric and Bernoulli variables Y and Z , we show that the first and second factorial moments of X are in one-to-one correspondence with the parameter pair (π, p) . This is a sufficient condition for identifiability. We have

$$EY = \frac{\pi}{1-\pi}, \quad EY(Y-1) = \frac{2\pi^2}{(1-\pi)^2}$$

and

$$EZ = p, \quad EZ(Z-1) = 0.$$

Let

$$m_1 = EX = \frac{\pi}{1-\pi} + p$$

and

$$m_2 = E(X(X-1))$$

$$= \frac{2\pi}{1-\pi} (p + \frac{\pi}{1-\pi}).$$

Then

$$m_2 = \frac{2\pi}{1-\pi} m_1$$

or

$$\pi = \frac{m_2}{2m_1 + m_2}$$

and thus

$$p = m_1 - \frac{m_2}{2m_1}.$$

Thus, the pairs (π, p) and (m_1, m_2) are in one-to-one correspondence.

Suppose a sample of size n is taken from a geometric-Bernoulli convolution. Let x_0, x_1, \dots, x_k be the k distinct integers observed, with frequencies n_0, n_1, \dots, n_k . We assume that $x_0 = 0$ and will treat the cases $n_0 = 0$ and $n_0 > 0$ separately. All other n_i are assumed positive with $\sum_{i=0}^k n_i = n$. The likelihood function for this sample may be written as

$$\begin{aligned}
 L(\underline{x}, \pi, p) &= \left(P(X=0) \right)^{n_0} \prod_{i=1}^k \left(P(X=x_i) \right)^{n_i} \\
 &= (1-p)^{n_0} (1-\pi)^{n_0} \prod_{i=1}^k \left((1-p)(1-\pi)\pi^{x_i} + p(1-\pi)\pi^{x_i-1} \right)^{n_i} \\
 &= (1-p)^{n_0} (1-\pi)^{n_0} \prod_{i=1}^k \pi^{n_i x_i - n_i} (1-\pi)^{n_i} ((1-p)\pi + p)^{n_i} \\
 &= (1-p)^{n_0} (1-\pi)^{n_0} \pi^{S-n+n_0} (1-\pi)^{n-n_0} ((1-p)\pi + p)^{n-n_0} \\
 &= (1-p)^{n_0} (1-\pi)^n \pi^{S-n+n_0} ((1-p)\pi + p)^{n-n_0}
 \end{aligned}$$

where $S = \sum_{i=1}^k n_i x_i$, that is, S is the sum of the n observations. The values of p and π for which $L > 0$ differ for different samples. We decompose the problem of maximum likelihood estimation into several mutually exclusive and exhaustive cases.

Case 1. $S = n - n_0$ (that is, all observations are either zero or one).

(a) $S = n - n_0$, $n_0 = n$. In this case, the likelihood is equal to

$$L = (1-p)^n (1-\pi)^n$$

which is positive for $\pi < 1$, $p < 1$, and is maximized over the unit square at $\hat{\pi} = 0$, $\hat{p} = 0$.

(b) $S = n - n_0$, $0 < n_0 < n$. Here,

$$L = (1-p)^{n_0} (1-\pi)^n \left((1-p)\pi + p \right)^{n-n_0}$$

is positive for $0 \leq \pi < 1$ and $-\frac{\pi}{1-\pi} < p < 1$.

Differentiating $\ln L$ one obtains

$$\frac{\partial}{\partial \pi} \ln L = -\frac{n}{1-\pi} + \frac{(n-n_0)(1-p)}{p + \pi(1-p)} = 0, \quad (2.1)$$

$$\frac{\partial}{\partial p} \ln L = -\frac{n_0}{1-p} + \frac{(n-n_0)(1-\pi)}{p + \pi(1-p)} = 0. \quad (2.2)$$

Solving (2.2) for p , we have

$$p = 1 - \frac{n_0}{n(1-\pi)}.$$

Substituting into (2.1), we obtain

$$\begin{aligned} 0 &= -\frac{n}{1-\pi} + \frac{(n-n_0)\left(\frac{n_0}{n(1-\pi)}\right)}{1 - \frac{n_0}{n(1-\pi)} + \pi\left(\frac{n_0}{n(1-\pi)}\right)} \\ &= -\frac{n}{1-\pi} + \frac{n_0(n-n_0)}{n(1-\pi) - n_0 + n_0\pi} \\ &= -\frac{n}{1-\pi} + \frac{n_0(n-n_0)}{(n-n_0)(1-\pi)} \\ &= \frac{n_0 - n}{1-\pi}. \end{aligned} \quad (2.3)$$

Since (2.3) has no solution, there are no critical points in this case.

Inspecting the boundary of the unit square, one easily obtains that

$(\hat{\pi}, \hat{p}) = (0, S/n)$ is the MLE.

(c) $S = n - n_0$, $n_0 = 0$. Here,

$$L = (1-\pi)^n (p + \pi(1-p))^n$$

which is positive for $0 \leq \pi < 1$, $- \frac{\pi}{1-\pi} < p \leq 1$. Differentiating $\ln L$ one obtains

$$\frac{\partial}{\partial \pi} \ln L = - \frac{n}{1-\pi} + \frac{n(1-p)}{p + \pi(1-p)} = 0$$

$$\frac{\partial}{\partial p} \ln L = \frac{n(1-\pi)}{p + \pi(1-p)} = 0,$$

which, by inspection, has no solution. Again, inspecting the boundary of the unit square, we find the MLE to be $(\hat{\pi}, \hat{p}) = (0, 1)$.

Case 2. $S > n - n_0$ (that is, at least one observation exceeds 1).

(a) $S > n - n_0$, $n_0 = 0$. The likelihood is given by

$$L = (1-\pi)^n \pi^{S-n} ((1-p)\pi + p)^n$$

which is positive for $0 < \pi < 1$ and $- \frac{\pi}{1-\pi} < p \leq 1$. Now

$$\frac{\partial}{\partial \pi} \ln L = - \frac{n}{1-\pi} + \frac{S-n}{\pi} + \frac{n(1-p)}{(1-p)\pi + p} = 0$$

$$\frac{\partial}{\partial p} \ln L = \frac{n(1-\pi)}{(1-p)\pi + p} = 0.$$

These equations have no solution. Inspection of the boundary of the unit square identifies $(\hat{\pi}, \hat{p}) = (\frac{S-n}{S}, 1)$ as the MLE.

(b) $s > n - n_0$, $0 < n_0 < n$. Here,

$$L = (1-\pi)^n (1-p)^{n_0} \pi^{s-n+n_0} ((1-p)\pi + p)^{n-n_0}$$

which is positive for $0 < \pi < 1$, $-\frac{\pi}{1-\pi} < p < 1$. Now

$$\frac{\partial}{\partial \pi} \ln L = -\frac{n}{1-\pi} + \frac{s-n+n_0}{\pi} + \frac{(n-n_0)(1-p)}{(1-p)\pi + p} = 0 \quad (2.4)$$

$$\frac{\partial}{\partial p} \ln L = -\frac{n_0}{1-p} + \frac{(n-n_0)(1-\pi)}{(1-p)\pi + p} = 0. \quad (2.5)$$

Solving (2.5) for p , we obtain

$$p = 1 - \frac{n_0}{n(1-\pi)}. \quad (2.6)$$

Substituting (2.6) in (2.4), we find

$$\hat{\pi} = \frac{s-n+n_0}{s} \quad (2.7)$$

and thus

$$\hat{p} = 1 - \frac{sn_0}{n(n-n_0)}. \quad (2.8)$$

We show that this pair $(\hat{\pi}, \hat{p})$ maximizes L , whether or not $(\hat{\pi}, \hat{p})$ is in the unit square, that is, whether or not $\hat{p} > 0$. The second partial derivatives, evaluated at $(\hat{\pi}, \hat{p})$, can be shown to be equal to

$$\frac{\partial^2}{\partial \pi^2} \ln L \Big|_{(\hat{\pi}, \hat{p})} = - \left[\frac{ns^2}{(n-n_0)^2} + \frac{s^2}{s-n+n_0} + \frac{s^2 n_0^2}{(n-n_0)^3} \right]$$

$$\frac{\partial^2}{\partial p^2} \ln L \Big|_{(\hat{\pi}, \hat{p})} = - \frac{n^3(n-n_0)}{s^2 n_0}$$

$$\frac{\partial^2}{\partial p \partial \pi} \ln L \Big|_{(\hat{\pi}, \hat{p})} = - \frac{n^2}{n-n_0} .$$

To show that the Hessian is negative definite at $(\hat{\pi}, \hat{p})$, we note that the diagonal terms are negative, and the determinant of the Hessian is

$$\begin{aligned} & \frac{n^4}{n_0(n-n_0)} + \frac{n^3(n-n_0)}{n_0(s-n+n_0)} + \frac{n_0 n^3}{(n-n_0)^2} - \frac{n^4}{(n-n_0)^2} \\ &= \frac{n^4}{n_0(n-n_0)} + \frac{n^3(n-n_0)}{n_0(s-n+n_0)} - \frac{n^3(n-n_0)}{(n-n_0)^2} \\ &= \frac{n^3}{n_0} + \frac{n^3(n-n_0)}{n_0(s-n+n_0)}, \end{aligned}$$

which is positive.

Now $\hat{p} < 0$ if and only if $s > n(n-n_0)/n_0$, and in this case, the solution $(\hat{\pi}, \hat{p})$ of the likelihood equations given by (2.7) and (2.8) is not the maximum likelihood estimate. Since there is no critical point in the unit square, we check the boundary to obtain the MLE $(\hat{\pi}, \hat{p}) = (\frac{s}{s+n}, 0)$. It is a useful coincidence that the inequality

$$s > \frac{n(n-n_0)}{n_0}$$

is equivalent to

$$\frac{S}{S+n} < \frac{S-n+n_0}{S} .$$

The MLE in Case 2 (b) may thus be written

$$(\hat{n}, \hat{p}) = \left(\min\left(\frac{S}{S+n}, \frac{S-n+n_0}{S}\right), \max\left(0, 1 - \frac{S n_0}{n(n-n_0)}\right) \right). \quad (2.9)$$

One may easily check that (\hat{n}, \hat{p}) given by (2.9) is in fact the MLE for samples containing at least one nonzero observation (that is, for all cases except 1 (a)), the MLE being $(\hat{n}, \hat{p}) = (0, 0)$ otherwise.

We turn our attention to the problem of maximum likelihood estimation for samples from the Poisson-Bernoulli convolution. We are unable to display a closed form solution for the MLE here, but we prove that any nontrivial solution (i.e., such that $\hat{p} \neq 0, 1$) of the likelihood equations satisfies the first moment condition

$$\hat{p} + \hat{\theta} = S/n,$$

that is, the condition that results from equating sample and population first moments. We are thus able to reduce our numerical search for the MLE to a one-dimensional problem.

Let X_1, \dots, X_n be a random sample from a Poisson-Bernoulli convolution. The likelihood function is given by

$$\begin{aligned}
 L(\tilde{x}, \theta, p) &= \prod_{i=1}^n \frac{e^{-\theta} \theta^{x_i-1}}{x_i!} ((1-p)\theta + px_i) \\
 &= \frac{e^{-n\theta} \theta^{S-n}}{\prod_{i=1}^n x_i!} \prod_{i=1}^n ((1-p)\theta + px_i).
 \end{aligned}$$

Thus, the likelihood equations are

$$\frac{\partial}{\partial \theta} \ln L = -n + \frac{S-n}{\theta} + (1-p) \sum_{i=1}^n \frac{1}{(1-p)\theta + px_i} = 0 \quad (2.10)$$

$$\frac{\partial}{\partial p} \ln L = \sum_{i=1}^n \frac{(x_i - \theta)}{(1-p)\theta + px_i} = 0. \quad (2.11)$$

By carrying out the indicated division, equation (2.11) may be rewritten as

$$\frac{n}{p} - (\theta + \frac{\theta}{p} (1-p)) \sum_{i=1}^n \frac{1}{(1-p)\theta + px_i} = 0 \quad (2.12)$$

provided $p \neq 0$. Assuming $p \neq 0, 1$, we substitute in for the sum in (2.12) an expression for this sum obtained from (2.10), yielding

$$\frac{n}{p} - (\theta + \frac{\theta}{p} (1-p)) (\frac{1}{1-p}) (n - \frac{S-n}{\theta}) = 0.$$

This is equivalent to

$$\frac{n}{p} - \frac{1}{1-p} (\frac{n\theta - S + n}{p}) = 0$$

or

$$\frac{n(1-p) - n\theta + S - n}{p(1-p)} = 0,$$

which reduces to the moment condition

$$p + \theta = S/n.$$

The moment condition established for solutions $(\hat{\theta}, \hat{p})$ of the likelihood equations is used in our numerical search for the MLE by substituting $\theta = S/n - p$ into equation (2.11), and searching for solutions of the equation

$$\sum_1^n \frac{(x_i - S/n + p)}{(S/n - p)(1-p) + px_i} = 0. \quad (2.13)$$

It can be shown that the point $(\hat{\theta}, \hat{p}) = (S/n, 0)$ is always a solution of the likelihood equations, so that our procedure for finding the MLE involves comparing the likelihood at this point with the likelihood at any point $(\hat{\theta}, \hat{p})$ for which $\hat{p} \in (0, 1]$ is a solution of (2.13).

III. RANDOM GENERATION AND NUMERICAL PROCEDURES

Random samples of size 50, 100 and 200 were generated from the geometric-Bernoulli and Poisson-Bernoulli convolutions using algorithms which employ random variables uniformly distributed on $[0, 1]$. Let u represent an observed uniform variable. Geometric random variables were obtained as follows:

1. Set $k = 0$.
2. Generate u .
3. If $u > \pi$, deliver k as $G(\pi)$; otherwise, set $k = k+1$ and go to 2.

Bernoulli random variables were generated by

1. Generate u .
2. If $u < p$, deliver $k = 1$ as $\theta(1,p)$; otherwise deliver $k = 0$.

Poisson random variables were generated from a sequence of uniform variables as follows:

1. Set $a = 1$, $k = 0$.
2. Generate u .
3. Let $a = a \cdot u$.
4. If $a \leq e^{-\theta}$, deliver k as $P(\theta)$; if $a > e^{-\theta}$, set $k = k+1$ and go to 2.

It is easy to show that the output k obtained by this algorithm has a Poisson distribution with parameter θ . The algorithm is based on a technique proposed by Ahrens, Dieter and Grube (1970).

The numerical procedure used to calculate the MLE for the parameters of the Poisson-Bernoulli convolution deserves some comment. Initially, the Newton-Raphson algorithm was used, but it was found to be inadequate. Several data sets gave rise to multiple roots, and there is no guarantee that the Newton-Raphson procedure converges to the MLE. The fact that $(\hat{\theta}, \hat{p}) = (S/n, 0)$ is always a solution of the likelihood equations seemed to cause difficulty. There were a number of examples in which NR converged to $(S/n, 0)$ when the MLE was elsewhere. This occurred in an example where the sample consisted of zeros and ones, in which case the MLE is $(\hat{\theta}, \hat{p}) = (0, S/n)$. The method of moments estimate of (θ, p) was used as an initial value for each sample.

After reducing the problem to one dimension, we opted for a search procedure which is guaranteed to find a root in $[0,1]$ if one exists. Interval halving and the method of false positions were considered. Barnett (1966) recommends the method of false positions, but makes incorrect claims concerning

this method. He claims incorrectly (see p. 159) that one can guarantee that the computed MLE is within an arbitrarily small distance of the true MLE. By partitioning an interval into subintervals of length ϵ , and then applying FP to each subinterval, one may find one root in each of several subintervals. Even if one of these roots is within ϵ of the true MLE it may happen that among all the roots found, the one at which the likelihood is largest is not within ϵ of the true MLE. Interval halving suffers from the same deficiency, of course, but was adopted because of its logical simplicity. The interval (.0001, .9999) was partitioned into ten intervals of equal length. The number ten was selected after partitioning the interval into 90 subintervals failed to produce more roots in any of 20 data sets on which we applied the procedure. A subinterval was identified as containing a root if the function in (2.13) took on different signs when evaluated at the endpoints of the interval. When the existence of a root was determined, the interval was halved, and the same criterion applied iteratively until the root was approximated to the desired level of accuracy. Once the roots uncovered by our procedure were well approximated, the likelihood was evaluated at each root and at the endpoints $p = 0$ and $p = 1$. The pair (θ, p) among these which maximized the likelihood was dubbed the MLE, and was used in our calculation of the likelihood ratio statistic.

IV. POWER ESTIMATION FOR χ^2 AND LIKELIHOOD RATIO TESTS

Five hundred samples of sizes 50, 100, 200 were taken from the two signal plus noise distributions under study at a variety of values of the parameter pair. The approximate χ^2 statistic appropriate for each test was tabulated and compared to the 5% cutoff of the χ^2 distribution with appropriate

degrees of freedom. In the case of the χ^2 goodness of fit test, the limiting null distribution of the χ^2 statistic when parameters are estimated by MLE's is given by Chernoff and Lehmann (1954). It is shown there that the limiting distribution is "between" χ_{k-1}^2 and χ_{k-r-1}^2 where k is the number of cells and r is the number of estimated parameters. In our simulation, the test which uses the 5% cutoff point of χ_{k-2}^2 was used throughout. This procedure slightly overestimates the power. For the likelihood ratio procedure, $-2 \ln \lambda$ was assumed to have χ_1^2 as its null distribution, where λ is the likelihood ratio statistic.

For each of the parameter pairs indicated, the proportion of rejections of the hypothesis $H_0: p = 0$ (postulating a geometric or a Poisson model for the data) in 500 repetitions was tabulated. The results are as follows:

TABLE I: $G(\pi) * \theta(1,p)$

$N = 50$

P	π	.35		.50		.65		.80	
		χ^2	LR	χ^2	LR	χ^2	LR	χ^2	LR
.00	.040	.040	.040	.048	.042	.044	.046	.056	.032
.10	.052	.086	.072	.066	.062	.076	.074	.068	
.20	.196	.254	.144	.208	.094	.140	.064	.112	
.30	.366	.472	.234	.372	.198	.320	.112	.202	
.40	.596	.760	.436	.610	.262	.500	.104	.324	
.50	.840	.902	.658	.860	.418	.710	.176	.422	
.60	.970	.998	.818	.942	.586	.870	.222	.674	
.70	.994	.998	.934	.988	.752	.966	.204	.798	
.80	.992	1.00	.934	1.00	.772	1.00	.250	.938	
.90	.840	1.00	.742	1.00	.588	1.00	.298	.994	
1.00	.578	1.00	.578	1.00	.484	1.00	.332	1.00	

N = 100

P π	.35		.50		.65		.80	
	χ^2	LR	χ^2	LR	χ^2	LR	χ^2	LR
.00	.042	.024	.044	.042	.054	.038	.086	.028
.10	.092	.150	.072	.116	.062	.074	.076	.084
.20	.300	.424	.224	.364	.154	.252	.100	.192
.30	.582	.770	.378	.636	.226	.508	.148	.340
.40	.882	.956	.708	.874	.428	.750	.196	.520
.50	.970	.992	.878	.984	.686	.922	.310	.744
.60	.998	1.00	.984	1.00	.890	.990	.438	.886
.70	1.00	1.00	1.00	1.00	.986	1.00	.640	.990
.80	1.00	1.00	1.00	1.00	.992	1.00	.688	1.00
.90	.992	1.00	.970	1.00	.898	1.00	.604	1.00
1.00	.900	1.00	.840	1.00	.730	1.00	.498	1.00

N = 200

P π	.35		.50		.65		.80	
	χ^2	LR	χ^2	LR	χ^2	LR	χ^2	LR
.00	.060	.036	.058	.030	.048	.028	.060	.022
.10	.128	.242	.104	.214	.056	.162	.060	.112
.20	.468	.702	.318	.574	.190	.468	.102	.300
.30	.882	.966	.676	.910	.410	.740	.166	.514
.40	.982	1.00	.922	.988	.720	.954	.324	.832
.50	1.00	1.00	.996	1.00	.938	1.00	.542	.956
.60					.998	1.00	.792	.994
.70							.946	1.00
.80	1.00	1.00	1.00	1.00	1.00	1.00	.990	1.00
.90					.998	1.00	.954	1.00
1.00	1.00	1.00	.990	1.00	.952	1.00	.774	1.00

TABLE II: $P(\theta) * \beta(1,p)$ $N = 50$

θ	p	$.50$		1.0		2.0		4.0	
		χ^2	LR	χ^2	LR	χ^2	LR	χ^2	LR
.00	Sample size too small for $\theta = .50$.032	.040	.040	.030	.048	.056
.10				.032	.022	.034	.042	.046	.042
.20				.048	.042	.042	.050	.044	.044
.30				.052	.068	.052	.054	.038	.032
.40				.064	.102	.052	.066	.046	.028
.50				.102	.172	.064	.088	.042	.040
.60				.120	.292	.060	.180	.040	.052
.70				.202	.474	.072	.192	.036	.100
.80				.270	.716	.124	.350	.048	.086
.90				.278	.948	.148	.562	.062	.130
1.00				.308	1.00	.176	.922	.064	.156

 $N = 100$

θ	p	$.50$		1.0		2.0		4.0	
		χ^2	LR	χ^2	LR	χ^2	LR	χ^2	LR
.00	Sample size too small for $\theta = .50$.052	.032	.046	.032	.056	.028	.044	.046
.10		.040	.032	.040	.026	.054	.046	.072	.040
.20		.058	.042	.056	.052	.030	.038	.050	.036
.30		.082	.146	.056	.092	.052	.058	.060	.054
.40		.166	.288	.064	.148	.056	.090	.050	.048
.50		.364	.602	.140	.320	.054	.114	.034	.088
.60		.654	.850	.242	.522	.074	.174	.058	.114
.70		.884	.978	.470	.794	.114	.364	.046	.206
.80		.990	.998	.718	.948	.136	.538	.078	.234
.90		.994	1.00	.832	1.00	.254	.816	.096	.362
1.00		.852	1.00	.582	1.00	.402	.998	.126	.448

θ	<u>N = 200</u>							
	.50		1.0		2.0		4.0	
P	χ^2	LR	χ^2	LR	χ^2	LR	χ^2	LR
.00	.044	.024	.052	.038	.066	.030	.046	.034
.10	.064	.040	.068	.044	.052	.024	.048	.050
.20	.072	.112	.040	.062	.048	.050	.042	.038
.30	.128	.234	.056	.132	.044	.056	.044	.054
.40	.348	.588	.110	.254	.050	.134	.046	.086
.50	.652	.868	.260	.514	.052	.202	.072	.108
.60	.952	.998	.490	.810	.120	.350	.046	.134
.70	1.00	1.00	.830	.976	.194	.570	.072	.262
.80	1.00	1.00	.972	.998	.372	.858	.076	.372
.90	1.00	1.00	.996	1.00	.538	.976	.112	.602
1.00	1.00	1.00	.898	1.00	.716	1.00	.178	.836

The estimates of the power of χ^2 and likelihood ratio tests tabulated above lead us to conclude that these tests have the following general characteristics. For a fixed value of the Bernoulli parameter, the power of both tests decrease as the geometric parameter π or the Poisson parameter θ increases. This agrees with our intuition in that as π or θ increases, the spread of the distribution increases, and it is correspondingly more difficult to detect a Bernoulli component. For fixed θ or π , the power of the likelihood ratio test increases with p . It is interesting to note that the χ^2 test does not behave similarly. The power curve seems to be parabolic in p for fixed π or θ , and the power is often seen to be higher at $p = .7$ or $.8$ than it is at $p = 1$. Insofar as the case $p = 1$ might be considered the most radical departure from the null hypothesis among the class of alternatives being considered, this feature of the χ^2 test is undesirable.

The estimated power curves of the likelihood ratio test generally (but not uniformly) dominate those of the χ^2 test. The χ^2 test dominates only for small values of p , and this is easily explained by the fact that the likelihood ratio test is somewhat conservative, achieving an actual significance level in the neighborhood of .03 compared to the nominal level of .05, while the χ^2 test is slightly anti-conservative. Thus, the power curves for the χ^2 test are generally above those for the LR test at $p = 0$, and the reversal takes place at some $p > 0$. The superiority of the LR test is occasionally striking. If one looks at the ratio of estimated power, one finds many examples of this ratio exceeding 3. For samples of size 50 generated from the geometric-Bernoulli convolution with $\pi = .8$ and $p = .7$, we were able to detect a Bernoulli component for only 20% of the samples using a χ^2 test, while with the LR test, it was detected in 80% of the samples. Similar comparisons can be made from our simulation of samples from the Poisson-Bernoulli convolution.

The class of signal plus noise distributions comprises a huge collection of probability models which contain the standard models as degenerate cases. There are many sources in nature of data which might reasonably be modeled by a signal plus noise distribution. The Geiger counter data set is a classic example -- a number of other examples are mentioned in Samaniego (1976). The power of the χ^2 goodness of fit test at signal plus noise alternatives tends to be low -- at least by comparison with other available tests. Approximate likelihood ratio tests, where maximum likelihood estimates of parameters of a signal plus noise distribution are obtained numerically, provide a means for checking the goodness of fit of standard probability models against a large class of alternatives. Our simulation results indicate

that the dividends from implementing such tests instead of the χ^2 test can be substantial.

REFERENCES

Ahrens, J. H., Dieter, U., and Grube, A. (1970). Pseudo random numbers - a new proposal for the choice of multiplicators. Computing 6, 121-38.

Barnett, V. D. (1966). Evaluation of the maximum likelihood estimator where the likelihood equation has multiple roots. Biometrika 53, 151-65.

Chernoff, H. and Lehmann, E. L. (1954). The use of maximum likelihood estimates in χ^2 tests of goodness of fit. Annals of Mathematical Statistics 25, 579-86.

Crow, E. L. and Bardwell, G. E. (1965). Estimation of the parameters of the hyper-Poisson distributions. G. P. Patil (editor), Classical and Contagious Discrete Distributions, Oxford: Pergamon Press, 127-40.

Samaniego, F. J. (1976). A characterization of convoluted Poisson distributions with applications to estimation. Journal of the American Statistical Association 71, 475-79.

Samaniego, F. J. (1977). A characterization of convoluted geometric distributions. Technical Report, Department of Mathematics, University of California, Davis.

Sclove, S. L. and Van Ryzin, J. (1969). Estimating the parameters of a convolution. Journal of the Royal Statistical Society, B 31, 181-91.

Sclove, S. L. (1977). Infinitely divisible distributions in statistical inference: heavy-tailed distributions and convolution models. Technical Report, Department of Mathematics, University of Illinois at Chicago Circle.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) AFOSR-TR-77-1256	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) ON THE POWER OF THE χ^2 SQUARED GOODNESS OF FIT TEST AT SIGNAL PLUS NOISE ALTERNATIVES.	5. TYPE OF REPORT & PERIOD COVERED (9) Interim rept.	
6. AUTHOR(s) (10) Lee D. Kaiser and Francisco J. Samaniego	7. CONTRACT OR GRANT NUMBER(s) (15) AFOSR-77-3180	8. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS University of California Department of Mathematics Davis, California 95616	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 61102F 2304 A5 (17) 12 A5	
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM Bolling AFB, DC 20332	12. REPORT DATE (11) September 1977	
13. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (14) TR-4 (12) 25p.	14. SECURITY CLASS. (of this report) UNCLASSIFIED	
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) χ^2 , likelihood ratio, maximum likelihood estimation, power, convolution. <i>chi square</i>		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The χ^2 goodness of fit test is widely used as a means of validation of a probability model. Acceptance of a probability model on the basis of a χ^2 test is a dangerous procedure in the absence of a thorough knowledge of the power of the test for a broad class of reasonable alternatives. The power of the χ^2 test		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract

test is examined via simulation at signal plus noise alternatives to the geometric and the Poisson distribution, and the power is seen to compare unfavorably with that of the likelihood ratio test. Likelihood ratio tests are thus advanced as goodness of fit criteria when signal plus noise alternatives are deemed relevant.

UNCLASSIFIED